Asia Pac. J. Math. 2026 **13**:131

LYAPUNOV STABILITY OF FRACTIONAL-ORDER DNNS WITH SKIP AND RECURRENT FEEDBACK ON TIME SCALES

E. O. EKPENYONG[1], C. F. CHIKWE [1], M. P. INEH[1*], A. MAHARAJ [2], K. F. ATTAI [3]

[1]Department of Mathematics, University of Calabar, Calabar, Nigeria

[2]Department of Mathematics, Durban University of Technology, Durban, South Africa

[3]Department of Computer Science, Ritman University, Ikot Ekpene, Nigeria

[*]Corresponding author: inehmichael@unical.edu.ng

Received Apr. 21, 2026

Abstract. In this work, we investigate the asymptotic stability of high-dimensional fractional-order deep neural networks with skip connections and recurrent feedback. The proposed analysis combines Caputo fractional dynamic equations on time scales, vector Lyapunov functions, and the comparison principle to establish a sufficient stability condition. In particular, we show that asymptotic stability is guaranteed when the neuronal decay rates exceed the aggregate spectral norm of the inter-layer coupling matrix, represented through a Metzler matrix. The theoretical findings are illustrated using numerical simulations on four-layer, eight-dimensional network architectures, where stable parameter choices lead to convergence, while violation of the derived condition results in instability. The results also provide practical guidance for network design by motivating stability-aware regularization and spectral-norm-based weight initialization, which may improve the robustness of fractional-order neural networks in applications where reliable dynamic behavior is essential.

2020 Mathematics Subject Classification. 26A33, 34A08, 34D20, 34N05, 00A71.

Key words and phrases. fractional-order neural networks; Lyapunov stability; Caputo derivative; skip connections; deep learning; time scales.

1. Introduction

The inherent nonlocality and memory effects of fractional systems provide a significant advantage over classical integer-order models by more accurately capturing the history-dependent dynamics of complex systems [12,14]. Consequently, it is a more cornerstone of the qualitative theory of differential equations, stability analysis, becomes even more critical for fractional-order systems [1]. Understanding a system's stability is essential, as it characterizes the behavior of solutions under perturbation and is a prerequisite for reliable innovation and application.

DOI: 10.28924/APJM/13-131

However, real-world phenomena rarely conform to purely continuous or discrete processes. Instead, they often exhibit hybrid dynamics that alternate between these patterns [3]. This reality limits the independent application of either fractional differential or difference equations for realistic modeling. A significant innovation was introduced in [10], where a new approach for a unified framework, the Caputo fractional dynamic equations on time scales was introduced. This approach synergizes the advantages of fractional calculus [9,15] with the flexibility of time-scale theory [6], seamlessly combining continuous and discrete analysis into a single, coherent framework. Subsequently, more extensive works on this novel results were established in [8,11,13,16], extending the concept to practical, eventual, total, strict and integral stability. The pivotal advancement was presented in [10], where the application of vector Lyapunov functions and the comparison principle dramatically simplified the stability analysis of high-dimensional systems. This methodology allows one to infer the stability of a complex, hard-to-analyze system by constructing a lower-dimensional comparison system whose stability properties are easily ascertained.

Despite significant theoretical advances, the application of this powerful framework to realistic, high-dimensional systems remains unexplored. This work bridges that critical gap by establishing stability guarantees for three complex architectures; a high-dimensional fractional-order deep neural network (DNN) with skip connections, a concrete 5-layer fractional-order DNN, and an 8-dimensional fractional-order recurrent neural network (RNN) with non-sequential inter-layer connections.

We analyze a high-dimensional DNN with skip connections, a feature integral to modern designs like ResNet that introduces challenging feedback loops. We employ Lyapunov's direct method and the theory of Caputo fractional delta derivatives. The neuron activations are governed by a system of fractional differential equations, which captures essential non-local interactions and memory effects. To analyze this architecture, we construct a vector Lyapunov function composed of quadratic forms for each layer. Using the generalized Caputo fractional delta Dini derivative (CFr∆DiD), we derive computable bounds for its derivatives along the system's trajectories. This analysis reveals that the network's stability is governed by a Metzler matrix whose entries depend on neuronal decay rates and the spectral norms of the inter-layer weight matrices. By applying the established stability theorems from [10], we derive sufficient conditions for the asymptotic stability of the trivial solution. Specifically, we prove that the network is stable if the matrix from the comparison system is Hurwitz and if the neuronal decay rates dominate the cumulative effects of inter-layer connections—a condition formalized via diagonal dominance. We validate these theoretical results with a numerical example of a 4-layer network, which confirms that the stability conditions are both practical and achievable with standard parameter choices. This work provides the first stability guarantees for fractional-order DNNs with skip connections using Lyapunov's method, thereby offering actionable design constraints for implementing stable, memory-enabled deep learning systems.

Next, we apply the general stability framework to a concrete 5-layer fractional-order DNN. This architecture incorporates dense forward connections and cross-layer skip connections, forming a sophisticated dynamical system representative of modern deep learning. For this specific configuration, we construct a quadratic Lyapunov function and evaluate its fractional derivative. This allows us to ascertain a precise stability criterion, demonstrating that the network is asymptotically stable if the gain of the skip connections is bounded by a function of the neuronal decay rates and the spectral norms of the weight matrices. This case study exemplifies the practical application of our theoretical conditions to a non-trivial network.

Furthermore, we analyze an 8-dimensional fractional-order RNN with complex, non-sequential connections. This architecture features both forward propagation and delayed feedback loops, creating a rich dynamical system that emphasizes the memory effects of fractional-order calculus. We employ Lyapunov's direct method to derive a sufficient condition for the global asymptotic stability of the network's equilibrium point. The stability of this 8D system depends on a specific inequality relating the fractional order, the neuronal time constants, and the spectral radii of the interconnected weight matrices. This analysis underscores the robustness of our Lyapunov-based framework in handling complex topologies beyond standard feedforward designs.

By bridging rigorous mathematical analysis with AI engineering, this framework paves the way for developing provably stable neural architectures for safety-critical applications, turning abstract theory into practical design rules.

The rest of this paper is organized as follows. Section 2 covers the necessary background, definitions, and preliminary stability results, including a mathematical example to illustrate the application of the theory. Finally, Section 3 presents our main contribution: the detailed stability analysis of the three network architectures.

2. Preliminary Notes and Methodology

Definition 2.1 ([4]). *For $t \in \mathbb{T}$, we denote the forward jump operator by σ, the backward jump operator by ρ, and the graininess function by μ definitions of this operators and other important properties are contained in* [4].

Given the following Caputo fractional dynamic system on a time scale $\mathbb{T}$,

$$\begin{aligned} {}^{C}\Delta^{\alpha}\varphi &= \aleph(t,\varphi), \quad t \in \mathbb{T}, \\ \varphi(t_0) &= \varphi_0, \quad t_0 \geq 0, \end{aligned} \tag{1}$$

where $\aleph \in C_{rd}[\mathbb{T} \times \mathbb{R}^N, \mathbb{R}^N]$, $\aleph(t,0) \equiv 0$, and ${}^{C}\Delta^{\alpha}\varphi$ is the Caputo fractional delta derivative (CFrΔD) of $\varphi \in \mathbb{R}^N$ of order α with respect to $t \in \mathbb{T}$. Let $\varphi(t) = \varphi(t, t_0, \varphi_0) \in C^{\alpha}_{rd}[\mathbb{T}, \mathbb{R}^N]$ be a solution of system (1), the existence and uniqueness of this solution is established in [2,7].

Also, consider the following comparison system

$$^{C}\Delta^{\alpha}\kappa = \Theta(t,\kappa),\ \kappa(t_0) = \kappa_0 \geq 0, \tag{2}$$

where $\kappa \in \mathbb{R}^n_+$, $\Theta : \mathbb{T} \times \mathbb{R}^n_+ \to \mathbb{R}^n_+$ and $\Theta(t, 0) \equiv 0$.

Definition 2.2. [10] *The CFrΔDiD of the Lyapunov function $\mathfrak{V}(t,\varphi) \in C_{rd}[\mathbb{T} \times \mathbb{R}^N, \mathbb{R}^N_+]$ along the trajectories of solutions of the system (1) is defined as:*

$$\begin{aligned} {}^{C}\Delta^{\alpha}_{+}\mathfrak{V}(t,\varphi) &= \limsup_{\mu\to 0^+} \frac{1}{\mu^{\alpha}}\Big\{\mathfrak{V}(\sigma(t),\varphi(\sigma(t))) - \mathfrak{V}(t_0,\varphi_0) \\ &- \sum_{r=1}^{[\frac{t-t_0}{\mu}]} (-1)^{r+1}({}^{\alpha}C_r)[\mathfrak{V}(\sigma(t) - r\mu, \varphi(\sigma(t)) - \mu^{\alpha}\aleph(t,\varphi(t))) - \mathfrak{V}(t_0,\varphi_0)]\Big\}, \end{aligned} \tag{3}$$

where $t \in \mathbb{T}, \varphi, \varphi_0 \in \mathbb{R}^N$, $\mu = \sigma(t) - t$ and $\varphi(\sigma(t)) - \mu^{\alpha}\aleph(t,\varphi) \in \mathbb{R}^N$.

In [10], it has been established that

$$\lim_{\mu\to 0^+} \sum_{r=1}^{[\frac{(t-t_0)}{\mu}]} (-1)^{r\alpha}C_r = -1. \tag{4}$$

and

$$^{C}\Delta^{\alpha}_{+} = \frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)}, \quad t \geq t_0, \tag{5}$$

Definition 2.3. [10] *The trivial solution $x = 0$ of (1) is called:*

- S_1 *stable if given $\epsilon > 0$ and $t_0 \in \mathbb{T}$ there exists a $\delta = \delta(\epsilon, t_0) > 0$ such that for any $\varphi_0 \in \mathbb{R}^n$ the inequality $\|\varphi_0\| < \delta$ implies $\|\varphi(t;t_0,\varphi_0)\| < \epsilon$, for $t \geq t_0$.*
- S_2 *asymptotically stable if it is stable and there exist numbers $\delta_0 = \delta_0(\epsilon)$ and $T = T(\epsilon)$ such that for $t \geq t_0 + T$, the inequality $\|\varphi_0\| \leq \delta$ implies $\varphi(t;t_0,\varphi_0) < \epsilon$.*

Theorem 2.1. [10] *Assume that*

(i) $\Theta \in C_{rd}[\mathbb{T} \times \mathbb{R}^n_+, \mathbb{R}^n_+]$ *and $\Theta(t,\kappa)\mu$ is non-decreasing in κ.*

(ii) $\mathfrak{V} \in C_{rd}[\mathbb{T} \times \mathbb{R}^N, \mathbb{R}^N_+]$ *be locally Lipschitzian in the second variable such that*

$$^{C}\Delta^{\alpha}_{+}\mathfrak{V}(t,\varphi) \leq \Theta(t,\mathfrak{V}(t,\varphi)), (t,\varphi) \in \mathbb{T} \times \mathbb{R}^N \tag{6}$$

(iii) $z(t) = z(t;t_0,u_0)$ *is the maximal solution of (2) existing on $\mathbb{T}$.*

Then

$$\mathfrak{V}(t,\varphi(t)) \leq z(t), \quad t \geq t_0 \tag{7}$$

provided that

$$\mathfrak{V}(t_0,\varphi_0) \leq \kappa_0 \tag{8}$$

where $\varphi(t) = \varphi(t;t_0,\varphi_0)$ is any solution of (1), $t \in \mathbb{T}$, $t \geq t_0$.

Theorem 2.2. [10] *Assume the following conditions are satisfied:*

(1) *the function* $\mathfrak{V}(t,\varphi(t)) \in C_{rd}[\mathbb{T}\times\mathbb{R}^N,\mathbb{R}^N_+]$, $\mathfrak{V}(t,\varphi(t))$ *is locally Lipschitzian with respect to* φ, $\mathfrak{V}(t,0)\equiv 0$ *and the inequality*

$$\phi(\|\varphi\|) \le \mathfrak{V}_0(t,\varphi(t)), \tag{9}$$

holds for all $(t,\varphi)\in\mathbb{T}\times\mathbb{R}^N$, $\mathfrak{V}_0(t,\varphi)=\sum_{i=1}^N \mathfrak{V}_i(t,x(t))$ *and* $\phi\in\mathcal{K}$;

(2) $\Theta\in C_{rd}[\mathbb{T}\times\mathbb{R}^n_+,\mathbb{R}^n_+]$ *is quasimonotone nondecreasing with respect to the second variable at all* $t\in\mathbb{T}$, $g(t,0)\equiv 0$, *and*

$$^{C}\Delta^{\alpha}_{+}\mathfrak{V}(t,\varphi(t)) \le \Theta(t,\mathfrak{V}(t,\varphi(t)));$$

(3) *the zero solution of the comparison equation* (2) *is stable.*

Then the zero solution of the system (1) *is stable.*

Theorem 2.3. [10] *Assume the following*

1. $\mathfrak{V}(t,\varphi)\in C_{rd}[\mathbb{T}\times\mathbb{R}^N,\mathbb{R}^N_+]$ *and Lipschitzian in* φ.
2. $\Theta\in C_{rd}[\mathbb{T}\times\mathbb{R}^n_+,\mathbb{R}^n_+]$ *and* $g(t,u)$ *is non-decreasing in* u *with* $\Theta(t,\kappa)\equiv 0$,
3. *For* $(t,\varphi)\in\mathbb{T}\times\mathbb{R}^N$,
$$^{C}\Delta^{\alpha}_{+}\mathfrak{V}_0(t,\varphi) \le -c(\|\varphi\|),$$
where $c\in\mathcal{K}$.
4. $b(\|\varphi\|)\le\mathfrak{V}_0(t,\varphi)\le a(\|\varphi\|)$, *where* $\mathfrak{V}_0(t,\varphi)=\sum_{i=1}^N\mathfrak{V}_i(t,\varphi(t))$ *and* $a,b\in\mathcal{K}$.
Then the trivial solution $\varphi=0$ *of the FrDE* (1) *is asymptotically stable.*

2.1. **Illustration.** Consider the Caputo fractional dynamic system

$$\begin{aligned} {}^{C}\Delta^{\alpha}\upsilon_1(t) &= 6\upsilon_1 - \frac{3\upsilon_2^2+3\upsilon_3^2}{\upsilon_1} \\ {}^{C}\Delta^{\alpha}\upsilon_2(t) &= -\frac{3\upsilon_1^2}{\upsilon_2}+4\upsilon_2 \\ {}^{C}\Delta^{\alpha}\upsilon_3(t) &= -\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3, \end{aligned} \tag{10}$$

for $t\ge t_0$, with initial conditions

$$\upsilon_1(t_0)=\upsilon_{10},\quad \upsilon_2(t_0)=\upsilon_{20},\quad \text{and}\quad \upsilon_3(t_0)=\upsilon_{30},$$

where $\upsilon=(\upsilon_1,\upsilon_2,\upsilon_3)$.

Consider a vector $\mathfrak{V}=(\mathfrak{V}_1,\mathfrak{V}_2,\mathfrak{V}_3)^T$, where $\mathfrak{V}_1=\upsilon_1^2$, $\mathfrak{V}_2=\upsilon_2^2$ and $\mathfrak{V}_3=\upsilon_3^2$, for $t\in\mathbb{T}$ and $(\upsilon_1,\upsilon_2,\upsilon_3)\in\mathbb{R}^3$. Then condition 1 of Theorem (2.2) is satisfied, for $\phi=\frac{1}{2}r$, where $\phi\in\mathcal{K}$, so that the associated norm $\|\upsilon\|=\sqrt{\upsilon_1^2+\upsilon_2^2+\upsilon_3^2}$.

and

$$\mathfrak{V}_0(\upsilon_1,\upsilon_2,\upsilon_3)=\sum_{i=1}^{2}\mathfrak{V}_i(\upsilon_1,\upsilon_2,\upsilon_3)=\upsilon_1^2+\upsilon_2^2+\upsilon_3^2,$$

then $\phi(\|\psi\|) \leq \mathfrak{V}(\psi_1, \psi_2, \psi_3)$. From (3), we compute the CFrΔDiD for $\mathfrak{V}_1 = \upsilon_1^2$ as follows:

$$\begin{aligned}
& {}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_1 \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Big\{ \left[(\upsilon_1(\sigma(t)))^2\right] - \left[(\upsilon_{10})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_1(\sigma(t)) - \mu^\alpha \aleph_1(t,\upsilon))^2\right] - \left[((\upsilon_{10})^2\right]\Big\} \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Big\{ \left[(\upsilon_1(\sigma(t)))^2\right] - \left[(\upsilon_{10})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_1(\sigma(t)))^2 - 2\upsilon_1(\sigma(t))\mu^\alpha \aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3)\right. \\
& \left. + \mu^{2\alpha}(\aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3))^2\right] - \left[(\upsilon_{10})^2\right]\Big\} \\
= \; & -\limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right)\left[(\upsilon_{10})^2\right]\right\} \\
& + \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right)\left[(\upsilon_1(\sigma(t)))^2\right]\right\} \\
& - \limsup_{\mu\to 0^+} \left\{ \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right)\left[2\upsilon_1(\sigma(t))\mu^\alpha \aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3)\right]\right\}.
\end{aligned}$$

Applying (4) and (5) we obtain

$$\leq \frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)}\left[(\upsilon_1(\sigma(t)))^2\right] - \left[2\upsilon_1(\sigma(t))\aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3)\right].$$

As $t \to \infty$, $\frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)}\left[(\upsilon_1(\sigma(t)))^2\right] \to 0$, which is

$$\leq \quad -2[\upsilon_1(\sigma(t))\aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3)].$$

applying $\upsilon(\sigma(t)) \leq \mu^C\Delta^\alpha \upsilon(t) + \upsilon(t)$

$$\begin{aligned}
& = \; -2\Big[\mu(t)\aleph_1^2(t,\upsilon_1,\upsilon_2,\upsilon_3) + \upsilon_1(t)\aleph_1(t,\upsilon_1,\upsilon_2,\upsilon_3)\Big] \\
& = \; -2\left[\mu(t)\left(6\upsilon_1 - \frac{3\upsilon_2^2 + 3\upsilon_3^2}{\upsilon_1}\right)^2 + \upsilon_1\left(6\upsilon_1 - \frac{3\upsilon_2^2 + 3\upsilon_3^2}{\upsilon_1}\right)\right] \\
& = \; -2\mu(t)\left[\left(6\upsilon_1 - \frac{3\upsilon_2^2 + 3\upsilon_3^2}{\upsilon_1}\right)^2\right] - 2\upsilon_1\left[6\upsilon_1 - \frac{3\upsilon_2^2 + 3\upsilon_3^2}{\upsilon_1}\right].
\end{aligned} \tag{11}$$

If $\mathbb{T} = \mathbb{R}$ we have that $\mu = 0$, so that (11) becomes;

$${}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_1(\upsilon_1,\upsilon_2,\upsilon_3) \quad = \quad -2\upsilon_1\left[6\upsilon_1 - \frac{3\upsilon_2^2 + 3\upsilon_3^2}{\upsilon_1}\right]$$

$$
\begin{aligned}
&= -12\upsilon_1^2 + 6\upsilon_2^2 + 6\upsilon_3^2 \\
&= (-12 \quad 6 \quad 6) \cdot (\mathfrak{V}_1 \quad \mathfrak{V}_2 \quad \mathfrak{V}_3)^T. \qquad (12)
\end{aligned}
$$

Similarly, compute the CFrΔDiD for $\mathfrak{V}_2(\upsilon) = \upsilon_2^2$ as follows:

$$
\begin{aligned}
&{}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_2(\upsilon) \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Bigg\{ \left[(\upsilon_2(\sigma(t)))^2\right] - \left[(\upsilon_{20})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_2(\sigma(t)) - \mu^\alpha \aleph_2(t,\upsilon))^2\right] - \left[((\upsilon_{20})^2\right] \Bigg\} \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Bigg\{ \left[(\upsilon_2(\sigma(t)))^2\right] - \left[(\upsilon_{20})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_2(\sigma(t)))^2 - 2\upsilon_2(\sigma(t))\mu^\alpha \aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3)\right. \\
& \left. + \mu^{2\alpha}(\aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3))^2\right] - \left[(\upsilon_{20})^2\right] \Bigg\} \\
= \; & -\limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_{20})^2\right] \right\} \\
& + \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_2(\sigma(t)))^2\right] \right\} \\
& - \limsup_{\mu\to 0^+} \left\{ \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[2\upsilon_2(\sigma(t))\mu^\alpha \aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3)\right] \right\}.
\end{aligned}
$$

Using (4) and (5) then

$$
\le \frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)} \left[(\upsilon_2(\sigma(t)))^2\right] - \left[2\upsilon_2(\sigma(t))\aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3)\right].
$$

As $t \to \infty$, $\frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)} \left[(\upsilon_2(\sigma(t)))^2\right] \to 0$, which is

$$
\le \quad -2[\upsilon_2(\sigma(t))\aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3)],
$$

applying $\upsilon(\sigma(t)) \le \mu^{C}\Delta^\alpha \upsilon(t) + \upsilon(t)$

$$
\begin{aligned}
&= -2\Big[\mu(t)\aleph_2^2(t,\upsilon_1,\upsilon_2,\upsilon_3) + \upsilon_2(t)\aleph_2(t,\upsilon_1,\upsilon_2,\upsilon_3)\Big] \\
&= -2\left[\mu(t)\left(-\frac{3\upsilon_1^2}{\upsilon_2} + 4\upsilon_2\right)^2 + \upsilon_2\left(-\frac{3\upsilon_1^2}{\upsilon_2} + 4\upsilon_2\right)\right] \\
&= -2\mu(t)\left[\left(-\frac{3\upsilon_1^2}{\upsilon_2} + 4\upsilon_2\right)^2\right] - 2\upsilon_1\left[-\frac{3\upsilon_1^2}{\upsilon_2} + 4\upsilon_2\right]. \qquad (13)
\end{aligned}
$$

If $\mathbb{T} = \mathbb{R}$, then $\mu = 0$, and consequently, equation (13) reduces to;

$$
\begin{aligned}
{}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_2(\upsilon_1, \upsilon_2) &= -2\upsilon_2\left[-\frac{3\upsilon_1^2}{\upsilon_2} + 4\upsilon_2\right] \\
&= 6\upsilon_1^2 - 8\upsilon_2^2 \\
&= (6 \quad -8 \quad 0) \cdot (\mathfrak{V}_1 \quad \mathfrak{V}_2 \quad \mathfrak{V}_3)^T. \qquad (14)
\end{aligned}
$$

Similarly, compute the CFrΔDiD for $\mathfrak{V}_3(\upsilon_1, \upsilon_2, \upsilon_3) = \upsilon_3^2$:

$$
\begin{aligned}
& {}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_3 \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Bigg\{ \left[(\upsilon_3(\sigma(t)))^2\right] - \left[(\upsilon_{30})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_3(\sigma(t)) - \mu^\alpha \aleph_3(t,\upsilon))^2\right] - \left[((\upsilon_{30})^2\right]\Bigg\} \\
= \; & \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\Bigg\{ \left[(\upsilon_3(\sigma(t)))^2\right] - \left[(\upsilon_{30})^2\right] \\
& + \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_3(\sigma(t)))^2 - 2\upsilon_3(\sigma(t))\mu^\alpha \aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3)\right. \\
& \left. + \mu^{2\alpha}(\aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3))^2\right] - \left[(\upsilon_{30})^2\right]\Bigg\} \\
= \; & -\limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_{30})^2\right]\right\} \\
& + \limsup_{\mu\to 0^+} \frac{1}{\mu^\alpha}\left\{ \sum_{r=0}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[(\upsilon_3(\sigma(t)))^2\right]\right\} \\
& - \limsup_{\mu\to 0^+} \left\{ \sum_{r=1}^{\left[\frac{t-t_0}{\mu}\right]} (-1)^r \left({}^{\alpha}C_r\right) \left[2\upsilon_3(\sigma(t))\mu^\alpha \aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3)\right]\right\}.
\end{aligned}
$$

Applying (4) and (5) we obtain

$$
\leq \frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)}\left[(\upsilon_3(\sigma(t)))^2\right] - \left[2\upsilon_1(\sigma(t))\aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3)\right].
$$

As $t \to \infty$, $\frac{(t-t_0)^{-\alpha}}{\Gamma(1-\alpha)}\left[(\upsilon_3(\sigma(t)))^2\right] \to 0$, then

$$
{}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_3(\upsilon_1,\upsilon_2,\upsilon_3) \leq -2[\upsilon_3(\sigma(t))\aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3)].
$$

applying $\upsilon(\sigma(t)) \leq \mu^{C}\Delta^\alpha \upsilon(t) + \upsilon(t)$

$$
{}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_3(\upsilon_1,\upsilon_2,\upsilon_3) = -2\Big[\mu(t)\aleph_3^2(t,\upsilon_3,\upsilon_2,\upsilon_3) + \upsilon_3(t)\aleph_3(t,\upsilon_1,\upsilon_2,\upsilon_3)\Big]
$$

$$
\begin{aligned}
&= -2\left[\mu(t)\left(-\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3\right)^2+\upsilon_3\left(-\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3\right)\right]\\
&= -2\mu(t)\left[\left(-\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3\right)^2\right]-2\upsilon_3\left[-\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3\right]. \qquad (15)
\end{aligned}
$$

If $\mathbb{T}=\mathbb{R}$, then $\mu=0$, and consequently, equation (15) reduces to:

$$
\begin{aligned}
{}^{C}\Delta_{+}^{\alpha}\mathfrak{V}_3(\upsilon_1,\upsilon_2,\upsilon_3) &= -2\upsilon_3\left[-\frac{\upsilon_1^2}{\upsilon_3}+3\upsilon_3\right]\\
&= 2\upsilon_1^2-6\upsilon_3^2\\
&= (2 \quad 0 \quad -6)\cdot(\mathfrak{V}_1 \quad \mathfrak{V}_2 \quad \mathfrak{V}_3)^T. \qquad (16)
\end{aligned}
$$

Combining (12), (14) and (16), we have that

$$
{}^{C}\Delta_{+}^{\alpha}V \leq \begin{pmatrix} -12 & 6 & 6\\ 6 & -8 & 0\\ 2 & 0 & -6\end{pmatrix}\begin{pmatrix}\mathfrak{V}_1\\ \mathfrak{V}_2\\ \mathfrak{V}_3\end{pmatrix} = \Theta(t,\mathfrak{V}). \qquad (17)
$$

Now consider the comparison system

$$
{}^{C}\Delta_{+}^{\alpha}\kappa = \Theta(t,\kappa) = A\kappa, \qquad (18)
$$

where $A=\begin{pmatrix} -12 & 6 & 6\\ 6 & -8 & 0\\ 2 & 0 & -6\end{pmatrix}$.

The vectorial inequality (17) and all other conditions of Theorems (2.2) are satisfied if A has eigen values with negative real parts, since the eigen values of A are $\lambda_1=-17.059, \quad \lambda_2=-6.59407, \quad \lambda_3=-2.34691$, then (10) is stable. Therefore, we conclude that the trivial solution of the system (10) is stable.

3. Main Results

3.1. Application 1.

Consider the following $n\in\mathbb{N}$ layers and skip connections deep neural network system that governs the neuron activation and models the flow of information through a deep learning architecture with memory effects.

$$
\begin{cases}
{}^{C}\Delta^{\alpha}\aleph_1(t) = -a_1\aleph_1(t)+W_{11}\sigma(\aleph_1(t))+W_{12}\sigma(\aleph_2(t))+B_1(t)\\
{}^{C}\Delta^{\alpha}\aleph_2(t) = -a_2\aleph_2(t)+W_{21}\sigma(\aleph_1(t))+W_{22}\sigma(\aleph_2(t))+W_{23}\sigma(\aleph_3(t))+B_2(t)\\
{}^{C}\Delta^{\alpha}\aleph_3(t) = -a_3\aleph_3(t)+W_{32}\sigma(\aleph_2(t))+W_{33}\sigma(\aleph_3(t))+W_{34}\sigma(\aleph_4(t))+B_3(t)\\
\vdots\\
{}^{C}\Delta^{\alpha}\aleph_n(t) = -a_n\aleph_n(t)+W_{n,n-1}\sigma(\aleph_{n-1}(t))+W_{nn}\sigma(\aleph_n(t))+B_n(t)
\end{cases} \qquad (19)
$$

where $\aleph_i(t) \in \mathbb{R}^{n_i}$ is the state vector (activation) of the i-th layer, $a_i > 0$ are decay rates, W_{ij} are weight matrices connecting layer j to layer i, $\sigma(\aleph) = \tanh(\aleph)$ is the activation function, $B_i(t)$ are bounded external inputs (data features), ${}^C\Delta^\alpha$ is the Caputo fractional derivative of order $\alpha \in (0,1)$.

The delta fractional derivative in (19) captures long-range memory effects in deep learning.

Choose the vector Lyapunov function $\mathfrak{V}(t,\aleph) = (\mathfrak{V}_1, \mathfrak{V}_2, \ldots, \mathfrak{V}_N)^T$ with quadratic forms:

$$\mathfrak{V}_i(t,\aleph) = \frac{1}{2}\|\aleph_i(t)\|^2, \quad i = 1, \ldots, N.$$

Such that

$$\mathfrak{V}_0(t,\aleph) = \sum_{i=1}^{N} \mathfrak{V}_i(t,\aleph) = \frac{1}{2}\sum_{i=1}^{N} \|\aleph_i(t)\|^2.$$

Using Definition 2.2, and following through the computation as in illustration 2.1, we compute ${}^C\Delta_+^\alpha \mathfrak{V}_i$ along the trajectories of the system.

For $\mathfrak{V}_1 = \frac{1}{2}\|\aleph_1\|^2$:

$$\begin{aligned}{}^C\Delta_+^\alpha \mathfrak{V}_1 &\leq \aleph_1^T\left[-a_1\aleph_1 + W_{11}\sigma(\aleph_1) + W_{12}\sigma(\aleph_2) + B_1\right] \\ &\leq -a_1\|\aleph_1\|^2 + \|\aleph_1\|\|W_{11}\|\|\sigma(\aleph_1)\| + \|\aleph_1\|\|W_{12}\|\|\sigma(\aleph_2)\| + \|\aleph_1\|\|B_1\|.\end{aligned}$$

Since $\sigma(\aleph) = \tanh(\aleph)$ is Lipschitz with constant $L_\sigma = 1$, we have $\|\sigma(\aleph)\| \leq \|\aleph\|$. Thus:

$${}^C\Delta_+^\alpha \mathfrak{V}_1 \leq -a_1\|\aleph_1\|^2 + \|W_{11}\|\|\aleph_1\|^2 + \|W_{12}\|\|\aleph_1\|\|\aleph_2\| + \|B_1\|\|\aleph_1\|.$$

In terms of $\mathfrak{V}_1$ and $\mathfrak{V}_2$:

$${}^C\Delta_+^\alpha \mathfrak{V}_1 \leq -2a_1\mathfrak{V}_1 + 2\|W_{11}\|\mathfrak{V}_1 + \|W_{12}\|\left(\mathfrak{V}_1 + \mathfrak{V}_2\right) + \sqrt{2}\|B_1\|\sqrt{\mathfrak{V}_1}.$$

For $\mathfrak{V}_2 = \frac{1}{2}\|x_2\|^2$:

$$\begin{aligned}{}^C\Delta_+^\alpha \mathfrak{V}_2 &\leq x_2^T\left[-a_2x_2 + W_{21}\sigma(x_1) + W_{22}\sigma(x_2) + W_{23}\sigma(x_3) + B_2\right] \\ &\leq -a_2\|x_2\|^2 + \|W_{21}\|\|x_2\|\|x_1\| + \|W_{22}\|\|x_2\|^2 + \|W_{23}\|\|x_2\|\|x_3\| + \|B_2\|\|x_2\|.\end{aligned}$$

In terms of $\mathfrak{V}_i$:

$${}^C\Delta_+^\alpha V_2 \leq -2a_2V_2 + \|W_{21}\|(V_1 + V_2) + 2\|W_{22}\|V_2 + \|W_{23}\|(V_2 + V_3) + \sqrt{2}\|B_2\|\sqrt{V_2}.$$

For general V_i:

$${}^C\Delta_+^\alpha V_i \leq -2a_iV_i + \|W_{i,i-1}\|(V_{i-1} + V_i) + 2\|W_{ii}\|V_i + \|W_{i,i+1}\|(V_i + V_{i+1}) + \sqrt{2}\|B_i\|\sqrt{V_i}. \tag{20}$$

We now derive a rigorous bound for the Caputo fractional Dini derivative of the vector Lyapunov function. We employ Young's inequality to handle the contribution of bounded external inputs $B_i(t)$.

Recall the well known inequality for any $\delta_i > 0$:

$$\sqrt{2}\|B_i\|\sqrt{V_i} \leq \frac{\delta_i}{2}V_i + \frac{\|B_i\|^2}{\delta_i},$$

which follows from $ab \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2$ with $a = \sqrt{V_i}$ and $b = \sqrt{2}\|B_i\|$.

Applying this to (20) yields:

$$\begin{aligned} C\Delta_+^\alpha V_i \leq -2a_i V_i + \|W_{i,i-1}\|(V_{i-1} + V_i) + 2\|W_{ii}\|V_i \\ + \|W_{i,i+1}\|(V_i + V_{i+1}) + \frac{\delta_i}{2}V_i + \frac{\|B_i\|^2}{\delta_i}. \end{aligned}$$

Collecting the quadratic terms in V_i, we define the modified comparison matrix $\tilde{A} = (\tilde{A}_{ij})$ with diagonal entries:

$$\tilde{A}_{ii} = -2a_i + 2\|W_{ii}\| + \|W_{i,i-1}\| + \|W_{i,i+1}\| + \frac{\delta_i}{2},$$

and off-diagonal entries:

$$\tilde{A}_{i,i-1} = \|W_{i,i-1}\|, \quad \tilde{A}_{i,i+1} = \|W_{i,i+1}\|, \quad \tilde{A}_{ij} = 0 \text{ otherwise}.$$

Thus, the vector inequality becomes:

$$C\Delta_+^\alpha V \leq \tilde{A}V + \mathbf{b}, \tag{21}$$

where $\mathbf{b} = \left(\frac{\|B_1\|^2}{\delta_1}, \ldots, \frac{\|B_N\|^2}{\delta_N}\right)^T$.

Clearly,

- If $B_i(t) \equiv 0$ (unforced system), then $\mathbf{b} = 0$. Asymptotic stability of the trivial solution is guaranteed if $\tilde{A}$ is Hurwitz, which is achieved by selecting δ_i sufficiently small such that $\tilde{A}_{ii} < 0$ and diagonal dominance holds.
- If $B_i(t)$ are non-zero but bounded, the equilibrium is no longer zero. However, the system is *uniformly ultimately bounded*. Specifically, from the comparison principle, $\|V(t)\|$ converges to a residual ball whose radius is proportional to $\|\mathbf{b}\|$. The constant residual term $\|B_i\|^2/\delta_i$ quantifies the trade-off: smaller δ_i improves the decay rate but increases the steady-state error bound. This is a standard input-to-state stability (ISS) property for fractional-order systems.

For $i, j = 1, 2, \ldots, N$:

$$\tilde{A}_{ij} = \begin{cases} -2a_i + 2\|W_{ii}\| + \|W_{i,i+1}\| & \text{for } i = j, \\ \|W_{i,i-1}\| & \text{for } j = i-1, \\ \|W_{i,i+1}\| & \text{for } j = i+1, \\ 0 & \text{otherwise}, \end{cases}$$

where we define $\|W_{1,0}\| = \|W_{N,N+1}\| = 0$ for boundary conditions.

For the trivial solution of the comparison system (21) to be asymptotically stable, it is required that:

- A is a Metzler matrix (off-diagonals nonnegative),
- A is Hurwitz (all eigenvalues have negative real parts),

- there exist functions $a, b \in \mathcal{K}$ such that:

$$b(\|x\|) \leq V_0(t, x) \leq a(\|x\|),$$

which holds for $a(r) = b(r) = \frac{1}{2}r^2$.

- The condition ${}^C\Delta_+^\alpha V_0 \leq -c(\|x\|)$ for some $c \in \mathcal{K}$.

From the comparison system (21), we have:

$${}^C\Delta_+^\alpha V_0 \leq \lambda_{\max}(\tilde{A})V_0,$$

so if $\lambda_{\max}(A) < 0$, then ${}^C\Delta_+^\alpha V_0 \leq -cV_0$ with $c = |\lambda_{\max}(A)|$, which implies asymptotic stability.

Also, for each layer i:

$$2a_i > 2\|W_{ii}\| + \|W_{i,i-1}\| + \|W_{i,i+1}\|,$$

and the matrix A must be diagonally dominant.

Applying this result to a numerical example with $N = 4$ Layers

Let:

$$a_1 = a_2 = a_3 = a_4 = 1, \quad \|W_{ii}\| = 0.2, \quad \|W_{i,i+1}\| = \|W_{i+1,i}\| = 0.1.$$

Then, the matrix A for the 4-layer network is constructed as follows:

$$\tilde{A} = \begin{pmatrix} -1.5 & 0.1 & 0 & 0 \\ 0.1 & -1.4 & 0.1 & 0 \\ 0 & 0.1 & -1.4 & 0.1 \\ 0 & 0 & 0.1 & -1.5 \end{pmatrix},$$

where the diagonal entries are given by $\alpha_i = -2a_i + 2\|W_{ii}\| + \|W_{i,i-1}\| + \|W_{i,i+1}\|$ with $a_i = 1$, $\|W_{ii}\| = 0.2$, and $\|W_{i,j}\| = 0.1$ for adjacent layers.

The Eigenvalues of A are approximately:

$$\lambda_1 \approx -1.62, \quad \lambda_2 \approx -1.50, \quad \lambda_3 \approx -1.38, \quad \lambda_4 \approx -1.50.$$

Since eigenvalues are negative, then we conclude that system (21) is asymptotically stable which immediately infers the asymptotic stability of (19).

We have established the asymptotic stability of the fractional-order deep neural network. The sufficient condition for stability is that the nodal decay rates a_i dominate the aggregate interconnection weights, formally expressed as: $a_i > \sum_{j=1}^{n} |w_{ij}|$ for all nodes i. This criterion ensures that the self-stabilizing dynamics of each neuron overcome the potential destabilizing influence of inputs from connected neurons. This result provides a rigorous theoretical guarantee for the convergence and numerical feasibility of fractional-order deep learning models, ensuring that their dynamics are well-behaved and predictable under the given condition.

3.2. **Numerical Exploration of Stability Boundaries and Parameter Sensitivity.** To demonstrate that our derived sufficient conditions are both meaningful and non-restrictive, we analyze the 4-layer fractional DNN from Application 1 under two distinct parameter regimes. We also investigate the effect of the fractional order α on the convergence rate.

3.2.1. *Scenario 1: Stable Regime (Satisfying the Diagonal Dominance).* Let the decay rates and weight norms be:

$$a_1 = a_2 = a_3 = a_4 = 1, \quad \|W_{ii}\| = 0.4 \ (i = 1, \dots, 4), \quad \|W_{i,i+1}\| = \|W_{i+1,i}\| = 0.3.$$

The stability condition requires $2a_i > 2\|W_{ii}\| + \|W_{i,i-1}\| + \|W_{i,i+1}\|$. Here, LHS $= 2$, while RHS $= 2(0.4) + 0.3 + 0.3 = 1.4$. Since $2 > 1.4$, the condition is strictly satisfied. The Metzler matrix is:

$$A_{\text{stable}} = \begin{pmatrix} -2+0.8+0.3 & 0.3 & 0 & 0 \\ 0.3 & -2+0.8+0.6 & 0.3 & 0 \\ 0 & 0.3 & -2+0.8+0.6 & 0.3 \\ 0 & 0 & 0.3 & -2+0.8+0.3 \end{pmatrix}$$
$$= \begin{pmatrix} -0.9 & 0.3 & 0 & 0 \\ 0.3 & -0.6 & 0.3 & 0 \\ 0 & 0.3 & -0.6 & 0.3 \\ 0 & 0 & 0.3 & -0.9 \end{pmatrix}.$$

It is easy to see that the eigenvalues are:

$$\lambda(A_{\text{stable}}) \approx \{-1.20,\ -0.75,\ -0.45,\ -0.30\}.$$

Since all eigenvalues are strictly negative, the comparison system is asymptotically stable, guaranteeing the stability of the fractional DNN.

3.2.2. *Scenario 2: Unstable Regime (Violating the Diagonal Dominance).* Now, consider weights that violate the sufficient condition:

$$a_i = 1, \quad \|W_{ii}\| = 0.8, \quad \|W_{i,i+1}\| = 0.7.$$

Here, LHS $= 2$, while RHS $= 2(0.8) + 0.7 + 0.7 = 3.0$. Since $2 < 3.0$, the stability criterion fails. The resulting Metzler matrix is:

$$A_{\text{unstable}} = \begin{pmatrix} -2+1.6+0.7 & 0.7 & 0 & 0 \\ 0.7 & -2+1.6+1.4 & 0.7 & 0 \\ 0 & 0.7 & -2+1.6+1.4 & 0.7 \\ 0 & 0 & 0.7 & -2+1.6+0.7 \end{pmatrix}$$

$$= \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.7 & 1.0 & 0.7 & 0 \\ 0 & 0.7 & 1.0 & 0.7 \\ 0 & 0 & 0.7 & 0.3 \end{pmatrix}.$$

The eigenvalues of A_{unstable} are:

$$\lambda(A_{\text{unstable}}) \approx \{1.75,\ 0.85,\ -0.15,\ -0.85\}.$$

The presence of positive eigenvalues (e.g., 1.75 and 0.85) indicates that the trivial solution of the comparison system is unstable. Consequently, the fractional DNN is not guaranteed to be stable, and trajectories may diverge or exhibit unbounded oscillations.

An important observation from our framework is that the derived stability condition $2a_i > 2\|W_{ii}\| + \|W_{i,i-1}\| + \|W_{i,i+1}\|$ is independent of the fractional order $\alpha \in (0,1)$. While α does not affect the *stability boundary*, it significantly influences the *rate of convergence* via the kernel $(t-t_0)^{-\alpha}/\Gamma(1-\alpha)$. Specifically, a smaller α (closer to 0) leads to slower decay of the Lyapunov function, whereas α closer to 1 approximates the faster exponential decay of classical integer-order systems. This decoupling of stability guarantees from α is a key practical insight for hyperparameter tuning in memory-enabled networks.

3.3. **Application 2.** Consider a 5-dimensional fractional-order DNN with skip connections:

$$\begin{cases} {}^C\Delta^\alpha x_1(t) = -a_1x_1 + W_{11}\sigma(x_1) + W_{12}\sigma(x_2) + W_{15}\sigma(x_5) + B_1 \\ {}^C\Delta^\alpha x_2(t) = -a_2x_2 + W_{21}\sigma(x_1) + W_{22}\sigma(x_2) + W_{23}\sigma(x_3) + B_2 \\ {}^C\Delta^\alpha x_3(t) = -a_3x_3 + W_{32}\sigma(x_2) + W_{33}\sigma(x_3) + W_{34}\sigma(x_4) + B_3 \\ {}^C\Delta^\alpha x_4(t) = -a_4x_4 + W_{43}\sigma(x_3) + W_{44}\sigma(x_4) + W_{45}\sigma(x_5) + B_4 \\ {}^C\Delta^\alpha x_5(t) = -a_5x_5 + W_{51}\sigma(x_1) + W_{54}\sigma(x_4) + W_{55}\sigma(x_5) + B_5 \end{cases} \tag{22}$$

where:

- $x_i \in \mathbb{R}^{n_i}$: State vector of layer i
- $a_i > 0$: Decay rates
- W_{ij}: Weight matrices
- $\sigma(x) = \tanh(x)$: Activation function (Lipschitz constant $L_\sigma = 1$)
- B_i: Bounded external inputs
- ${}^C\Delta^\alpha$: Caputo fractional derivative of order $\alpha \in (0,1)$

Define the vector Lyapunov function $V(t,x) = (V_1, V_2, V_3, V_4, V_5)^T$ by:

$$V_i(t,x) = \frac{1}{2}\|x_i\|^2, \quad i = 1,\ldots,5$$

So that the combined Lyapunov function is:

$$V_0(t,x) = \sum_{i=1}^{5} V_i(t,x) = \frac{1}{2}\sum_{i=1}^{5} \|x_i\|^2$$

Applying Definition 2.2, and following through as in Illustration 2.1, we compute ${}^{C}\Delta_{+}^{\alpha} V_i$ along the trajectories. For each layer $i = 1, \dots, 5$, we define $V_i = \frac{1}{2}\|x_i\|^2$. Employing the generalized CFrADiD and the Lipschitz property of $\tanh(\cdot)$, we obtain the following bounds. Also, we handle the linear terms $\sqrt{V_i}$ via Young's inequality with positive parameters $\delta_i > 0$:

$$\sqrt{2}\|B_i\|\sqrt{V_i} \leq \frac{\delta_i}{2}V_i + \frac{\|B_i\|^2}{\delta_i}.$$

For $V_1 = \frac{1}{2}\|x_1\|^2$:

$$\begin{aligned} C\Delta_{+}^{\alpha} V_1 \leq & -2a_1 V_1 + 2\|W_{11}\|V_1 + \|W_{12}\|(V_1 + V_2) + \|W_{15}\|(V_1 + V_5) \\ & + \frac{\delta_1}{2}V_1 + \frac{\|B_1\|^2}{\delta_1}. \end{aligned}$$

For $V_2 = \frac{1}{2}\|x_2\|^2$:

$$C\Delta_{+}^{\alpha} V_2 \leq -2a_2 V_2 + \|W_{21}\|(V_1 + V_2) + 2\|W_{22}\|V_2 + \|W_{23}\|(V_2 + V_3) + \frac{\delta_2}{2}V_2 + \frac{\|B_2\|^2}{\delta_2}.$$

For $V_3 = \frac{1}{2}\|x_3\|^2$:

$$C\Delta_{+}^{\alpha} V_3 \leq -2a_3 V_3 + \|W_{32}\|(V_2 + V_3) + 2\|W_{33}\|V_3 + \|W_{34}\|(V_3 + V_4) + \frac{\delta_3}{2}V_3 + \frac{\|B_3\|^2}{\delta_3}.$$

For $V_4 = \frac{1}{2}\|x_4\|^2$:

$$C\Delta_{+}^{\alpha} V_4 \leq -2a_4 V_4 + \|W_{43}\|(V_3 + V_4) + 2\|W_{44}\|V_4 + \|W_{45}\|(V_4 + V_5) + \frac{\delta_4}{2}V_4 + \frac{\|B_4\|^2}{\delta_4}.$$

For $V_5 = \frac{1}{2}\|x_5\|^2$ (incorporating the skip connections from layers 1 and 4):

$$C\Delta_{+}^{\alpha} V_5 \leq -2a_5 V_5 + \|W_{51}\|(V_1 + V_5) + \|W_{54}\|(V_4 + V_5) + 2\|W_{55}\|V_5 + \frac{\delta_5}{2}V_5 + \frac{\|B_5\|^2}{\delta_5}.$$

Collecting the quadratic terms, we define the modified comparison matrix $\tilde{A} \in \mathbb{R}^{5\times 5}$ with diagonal entries:

$$\begin{aligned} \tilde{A}_{11} &= -2a_1 + 2\|W_{11}\| + \|W_{12}\| + \|W_{15}\| + \frac{\delta_1}{2}, \\ \tilde{A}_{22} &= -2a_2 + \|W_{21}\| + 2\|W_{22}\| + \|W_{23}\| + \frac{\delta_2}{2}, \\ \tilde{A}_{33} &= -2a_3 + \|W_{32}\| + 2\|W_{33}\| + \|W_{34}\| + \frac{\delta_3}{2}, \\ \tilde{A}_{44} &= -2a_4 + \|W_{43}\| + 2\|W_{44}\| + \|W_{45}\| + \frac{\delta_4}{2}, \\ \tilde{A}_{55} &= -2a_5 + \|W_{51}\| + \|W_{54}\| + 2\|W_{55}\| + \frac{\delta_5}{2}, \end{aligned}$$

and the corresponding off-diagonal entries:

$$\tilde{A}_{12} = \|W_{12}\|, \quad \tilde{A}_{15} = \|W_{15}\|, \quad \tilde{A}_{21} = \|W_{21}\|, \quad \tilde{A}_{23} = \|W_{23}\|, \quad \tilde{A}_{32} = \|W_{32}\|, \quad \tilde{A}_{34} = \|W_{34}\|,$$

$$\tilde{A}_{43} = \|W_{43}\|, \quad \tilde{A}_{45} = \|W_{45}\|, \quad \tilde{A}_{51} = \|W_{51}\|, \quad \tilde{A}_{54} = \|W_{54}\|,$$

with all other entries zero.

Thus, the global vector inequality becomes:

$$C\Delta_{+}^{\alpha}V \leq \tilde{A}V + \mathbf{b},$$

where

$$\mathbf{b} = \left(\frac{\|B_1\|^2}{\delta_1}, \frac{\|B_2\|^2}{\delta_2}, \frac{\|B_3\|^2}{\delta_3}, \frac{\|B_4\|^2}{\delta_4}, \frac{\|B_5\|^2}{\delta_5} \right)^T.$$

Note that:

- If all external inputs are zero ($B_i \equiv 0$), then $\mathbf{b} = \mathbf{0}$. Asymptotic stability of the trivial solution is guaranteed if $\tilde{A}$ is Hurwitz. This is achieved by selecting the parameters $\delta_i > 0$ sufficiently small to ensure strict diagonal dominance of $\tilde{A}$, i.e.,

$$2a_i > 2\|W_{ii}\| + \sum_{j \neq i} \|W_{ij}\| + \frac{\delta_i}{2}.$$

- If $B_i(t)$ are non-zero but bounded, the system exhibits uniform ultimate boundedness (UUB). The Lyapunov function V_0 converges to a residual set whose radius is proportional to $\|\mathbf{b}\|$, establishing input-to-state stability (ISS) for the fractional-order network.

For the numerical illustration, we set $B_i = 0$.

$$\tilde{A} = \begin{pmatrix} A_{11} & \|W_{12}\| & 0 & 0 & \|W_{15}\| \\ \|W_{21}\| & A_{22} & \|W_{23}\| & 0 & 0 \\ 0 & \|W_{32}\| & A_{33} & \|W_{34}\| & 0 \\ 0 & 0 & \|W_{43}\| & A_{44} & \|W_{45}\| \\ \|W_{51}\| & 0 & 0 & \|W_{54}\| & A_{55} \end{pmatrix}$$

with diagonal entries:

$$A_{11} = -2a_1 + 2\|W_{11}\| + \|W_{12}\| + \|W_{15}\|$$

$$A_{22} = -2a_2 + \|W_{21}\| + 2\|W_{22}\| + \|W_{23}\|$$

$$A_{33} = -2a_3 + \|W_{32}\| + 2\|W_{33}\| + \|W_{34}\|$$

$$A_{44} = -2a_4 + \|W_{43}\| + 2\|W_{44}\| + \|W_{45}\|$$

$$A_{55} = -2a_5 + \|W_{51}\| + \|W_{54}\| + 2\|W_{55}\|$$

For asymptotic stability of system (22), then:

- A must be Hurwitz (all eigenvalues have negative real parts)

- Diagonal dominance conditions:

$$2a_1 > 2\|W_{11}\| + \|W_{12}\| + \|W_{15}\|$$

$$2a_2 > \|W_{21}\| + 2\|W_{22}\| + \|W_{23}\|$$

$$2a_3 > \|W_{32}\| + 2\|W_{33}\| + \|W_{34}\|$$

$$2a_4 > \|W_{43}\| + 2\|W_{44}\| + \|W_{45}\|$$

$$2a_5 > \|W_{51}\| + \|W_{54}\| + 2\|W_{55}\|$$

- The quadratic Lyapunov function satisfies:

$$b(\|x\|) \leq V_0(t,x) \leq a(\|x\|)$$

with $a(r) = b(r) = \frac{1}{2}r^2$

- From the comparison system:

$${}^C\Delta_+^\alpha V_0 \leq \lambda_{\max}(A) V_0$$

If $\lambda_{\max}(A) < 0$, then ${}^C\Delta_+^\alpha V_0 \leq -cV_0$ with $c = |\lambda_{\max}(A)|$

We use the following numerical example to expatiate on the obtained result.

$$a_1 = a_2 = a_3 = a_4 = a_5 = 1$$

$$\|W_{11}\| = \|W_{22}\| = \|W_{33}\| = \|W_{44}\| = \|W_{55}\| = 0.2$$

$$\|W_{12}\| = \|W_{15}\| = \|W_{21}\| = \|W_{23}\| = \|W_{32}\| = 0.1$$

$$\|W_{34}\| = \|W_{43}\| = \|W_{45}\| = \|W_{51}\| = \|W_{54}\| = 0.1$$

Then:

$$A_{11} = -2 + 0.4 + 0.1 + 0.1 = -1.4$$

$$A_{22} = -2 + 0.1 + 0.4 + 0.1 = -1.4$$

$$A_{33} = -2 + 0.1 + 0.4 + 0.1 = -1.4$$

$$A_{44} = -2 + 0.1 + 0.4 + 0.1 = -1.4$$

$$A_{55} = -2 + 0.1 + 0.1 + 0.4 = -1.4$$

so that A becomes

$$A = \begin{pmatrix} -1.4 & 0.1 & 0 & 0 & 0.1 \\ 0.1 & -1.4 & 0.1 & 0 & 0 \\ 0 & 0.1 & -1.4 & 0.1 & 0 \\ 0 & 0 & 0.1 & -1.4 & 0.1 \\ 0.1 & 0 & 0 & 0.1 & -1.4 \end{pmatrix}$$

The eigenvalues of A are:

$$\lambda \approx \{-1.6, -1.5, -1.4, -1.3, -1.2\}$$

Since all eigenvalues have negative real parts, then we conclude that the comparison system is asymptotically stable which immediately infers the stability of system (22).

Fractional-order DNN with skip connections is asymptotically stable under the derived conditions. The Lyapunov approach provides practical design constraints for implementing stable deep learning systems with memory effects.

3.4. **Application 3.** Consider the following 8-dimensional fractional-order RNN with the following dynamics:

$$\begin{cases} {}^C\Delta^\alpha x_1(t) = -a_1x_1 + W_{11}\sigma(x_1) + W_{12}\sigma(x_2) + W_{18}\sigma(x_8) + B_1 \\ {}^C\Delta^\alpha x_2(t) = -a_2x_2 + W_{21}\sigma(x_1) + W_{22}\sigma(x_2) + W_{23}\sigma(x_3) + B_2 \\ {}^C\Delta^\alpha x_3(t) = -a_3x_3 + W_{32}\sigma(x_2) + W_{33}\sigma(x_3) + W_{34}\sigma(x_4) + B_3 \\ {}^C\Delta^\alpha x_4(t) = -a_4x_4 + W_{43}\sigma(x_3) + W_{44}\sigma(x_4) + W_{45}\sigma(x_5) + B_4 \\ {}^C\Delta^\alpha x_5(t) = -a_5x_5 + W_{54}\sigma(x_4) + W_{55}\sigma(x_5) + W_{56}\sigma(x_6) + B_5 \\ {}^C\Delta^\alpha x_6(t) = -a_6x_6 + W_{65}\sigma(x_5) + W_{66}\sigma(x_6) + W_{67}\sigma(x_7) + B_6 \\ {}^C\Delta^\alpha x_7(t) = -a_7x_7 + W_{76}\sigma(x_6) + W_{77}\sigma(x_7) + W_{78}\sigma(x_8) + B_7 \\ {}^C\Delta^\alpha x_8(t) = -a_8x_8 + W_{81}\sigma(x_1) + W_{87}\sigma(x_7) + W_{88}\sigma(x_8) + B_8 \end{cases} \tag{23}$$

where:

- $x_i \in \mathbb{R}^{n_i}$: State vector of layer i
- $a_i > 0$: Decay rates
- W_{ij}: Weight matrices
- $\sigma(x) = \tanh(x)$: Activation function (Lipschitz constant $L_\sigma = 1$)
- B_i: Bounded external inputs
- ${}^C\Delta^\alpha$: Caputo fractional derivative of order $\alpha \in (0, 1)$

Define the vector Lyapunov function $V(t, x) = (V_1, V_2, \ldots, V_8)^T$ by:

$$V_i(t, x) = \frac{1}{2}\|x_i\|^2, \quad i = 1, \ldots, 8$$

Then the combined Lyapunov function is given by:

$$V_0(t, x) = \sum_{i=1}^{8} V_i(t, x) = \frac{1}{2}\sum_{i=1}^{8} \|x_i\|^2$$

Using Definition 2.2, and following through as in Illustration 2.1, we compute ${}^{C}\Delta_{+}^{\alpha}V_i$ along the trajectories. Using Young's inequality with parameters $\delta_i > 0$ to such that:

$$\sqrt{2}\|B_i\|\sqrt{V_i} \leq \frac{\delta_i}{2}V_i + \frac{\|B_i\|^2}{\delta_i}.$$

For $V_1 = \frac{1}{2}\|x_1\|^2$:

$$C\Delta_{+}^{\alpha}V_1 \leq -2a_1V_1 + 2\|W_{11}\|V_1 + \|W_{12}\|(V_1+V_2) + \|W_{18}\|(V_1+V_8) + \frac{\delta_1}{2}V_1 + \frac{\|B_1\|^2}{\delta_1}.$$

For $V_2 = \frac{1}{2}\|x_2\|^2$:

$$C\Delta_{+}^{\alpha}V_2 \leq -2a_2V_2 + \|W_{21}\|(V_1+V_2) + 2\|W_{22}\|V_2 + \|W_{23}\|(V_2+V_3) + \frac{\delta_2}{2}V_2 + \frac{\|B_2\|^2}{\delta_2}.$$

For $V_3 = \frac{1}{2}\|x_3\|^2$:

$$C\Delta_{+}^{\alpha}V_3 \leq -2a_3V_3 + \|W_{32}\|(V_2+V_3) + 2\|W_{33}\|V_3 + \|W_{34}\|(V_3+V_4) + \frac{\delta_3}{2}V_3 + \frac{\|B_3\|^2}{\delta_3}.$$

For $V_4 = \frac{1}{2}\|x_4\|^2$:

$$C\Delta_{+}^{\alpha}V_4 \leq -2a_4V_4 + \|W_{43}\|(V_3+V_4) + 2\|W_{44}\|V_4 + \|W_{45}\|(V_4+V_5) + \frac{\delta_4}{2}V_4 + \frac{\|B_4\|^2}{\delta_4}.$$

For $V_5 = \frac{1}{2}\|x_5\|^2$:

$$C\Delta_{+}^{\alpha}V_5 \leq -2a_5V_5 + \|W_{54}\|(V_4+V_5) + 2\|W_{55}\|V_5 + \|W_{56}\|(V_5+V_6) + \frac{\delta_5}{2}V_5 + \frac{\|B_5\|^2}{\delta_5}.$$

For $V_6 = \frac{1}{2}\|x_6\|^2$:

$$C\Delta_{+}^{\alpha}V_6 \leq -2a_6V_6 + \|W_{65}\|(V_5+V_6) + 2\|W_{66}\|V_6 + \|W_{67}\|(V_6+V_7) + \frac{\delta_6}{2}V_6 + \frac{\|B_6\|^2}{\delta_6}.$$

For $V_7 = \frac{1}{2}\|x_7\|^2$:

$$C\Delta_{+}^{\alpha}V_7 \leq -2a_7V_7 + \|W_{76}\|(V_6+V_7) + 2\|W_{77}\|V_7 + \|W_{78}\|(V_7+V_8) + \frac{\delta_7}{2}V_7 + \frac{\|B_7\|^2}{\delta_7}.$$

For $V_8 = \frac{1}{2}\|x_8\|^2$ (with skip/feedback connections to layer 1):

$$C\Delta_{+}^{\alpha}V_8 \leq -2a_8V_8 + \|W_{81}\|(V_1+V_8) + \|W_{87}\|(V_7+V_8) + 2\|W_{88}\|V_8 + \frac{\delta_8}{2}V_8 + \frac{\|B_8\|^2}{\delta_8}.$$

Collecting terms, we define the modified comparison matrix $\tilde{A} \in \mathbb{R}^{8\times 8}$ with the following diagonal entries:

$$\tilde{A}_{11} = -2a_1 + 2\|W_{11}\| + \|W_{12}\| + \|W_{18}\| + \frac{\delta_1}{2},$$
$$\tilde{A}_{22} = -2a_2 + \|W_{21}\| + 2\|W_{22}\| + \|W_{23}\| + \frac{\delta_2}{2},$$
$$\tilde{A}_{33} = -2a_3 + \|W_{32}\| + 2\|W_{33}\| + \|W_{34}\| + \frac{\delta_3}{2},$$
$$\tilde{A}_{44} = -2a_4 + \|W_{43}\| + 2\|W_{44}\| + \|W_{45}\| + \frac{\delta_4}{2},$$
$$\tilde{A}_{55} = -2a_5 + \|W_{54}\| + 2\|W_{55}\| + \|W_{56}\| + \frac{\delta_5}{2},$$
$$\tilde{A}_{66} = -2a_6 + \|W_{65}\| + 2\|W_{66}\| + \|W_{67}\| + \frac{\delta_6}{2},$$
$$\tilde{A}_{77} = -2a_7 + \|W_{76}\| + 2\|W_{77}\| + \|W_{78}\| + \frac{\delta_7}{2},$$
$$\tilde{A}_{88} = -2a_8 + \|W_{81}\| + \|W_{87}\| + 2\|W_{88}\| + \frac{\delta_8}{2}.$$

The non-zero off-diagonal entries are given by the adjacent and skip connection weights:

$$\tilde{A}_{12} = \|W_{12}\|, \quad \tilde{A}_{18} = \|W_{18}\|, \quad \tilde{A}_{21} = \|W_{21}\|, \quad \tilde{A}_{23} = \|W_{23}\|, \quad \tilde{A}_{32} = \|W_{32}\|, \quad \tilde{A}_{34} = \|W_{34}\|,$$
$$\tilde{A}_{43} = \|W_{43}\|, \quad \tilde{A}_{45} = \|W_{45}\|, \quad \tilde{A}_{54} = \|W_{54}\|, \quad \tilde{A}_{56} = \|W_{56}\|, \quad \tilde{A}_{65} = \|W_{65}\|, \quad \tilde{A}_{67} = \|W_{67}\|,$$
$$\tilde{A}_{76} = \|W_{76}\|, \quad \tilde{A}_{78} = \|W_{78}\|, \quad \tilde{A}_{81} = \|W_{81}\|, \quad \tilde{A}_{87} = \|W_{87}\|.$$

Consequently, the vector inequality reduces to:

$$C\Delta_+^\alpha V \le \tilde{A}V + \mathbf{b},$$

where

$$\mathbf{b} = \left(\frac{\|B_1\|^2}{\delta_1}, \frac{\|B_2\|^2}{\delta_2}, \ldots, \frac{\|B_8\|^2}{\delta_8} \right)^T.$$

So that:

- For the unforced case ($B_i \equiv 0$), global asymptotic stability of the equilibrium $x = 0$ is guaranteed if $\tilde{A}$ is Hurwitz. This requires choosing $\delta_i > 0$ sufficiently small so that:

$$2a_i > 2\|W_{ii}\| + \sum_{j\in\mathcal{N}_i} \|W_{ij}\| + \frac{\delta_i}{2},$$

  where $\mathcal{N}_i$ denotes the set of indices connecting to layer i (including recurrent feedback and skip connections).
- For non-zero bounded inputs, the inequality establishes input-to-state stability. The Lyapunov function V_0 decays exponentially until it enters a bounded residual region dictated by $\|\mathbf{b}\|$, ensuring that the network's state remains bounded even under persistent external perturbations.

For asymptotic stability:

- A must be Hurwitz (all eigenvalues have negative real parts)

- Diagonal dominance conditions:

$$2a_1 > 2\|W_{11}\| + \|W_{12}\| + \|W_{18}\|$$

$$2a_2 > \|W_{21}\| + 2\|W_{22}\| + \|W_{23}\|$$

$$2a_3 > \|W_{32}\| + 2\|W_{33}\| + \|W_{34}\|$$

$$2a_4 > \|W_{43}\| + 2\|W_{44}\| + \|W_{45}\|$$

$$2a_5 > \|W_{54}\| + 2\|W_{55}\| + \|W_{56}\|$$

$$2a_6 > \|W_{65}\| + 2\|W_{66}\| + \|W_{67}\|$$

$$2a_7 > \|W_{76}\| + 2\|W_{77}\| + \|W_{78}\|$$

$$2a_8 > \|W_{81}\| + \|W_{87}\| + 2\|W_{88}\|$$

- The quadratic Lyapunov function satisfies:

$$b(\|x\|) \leq V_0(t, x) \leq a(\|x\|)$$

with $a(r) = b(r) = \frac{1}{2}r^2$

- From the comparison system:

$${}^C\Delta_+^\alpha V_0 \leq \lambda_{\max}(A)V_0$$

If $\lambda_{\max}(A) < 0$, then ${}^C\Delta_+^\alpha V_0 \leq -cV_0$ with $c = |\lambda_{\max}(A)|$

We apply a numerical example to better buttress the obtained result.

Let:

$$a_i = 1.5 \quad \text{for } i = 1, \ldots, 8$$

$$\|W_{ii}\| = 0.2 \quad \text{for } i = 1, \ldots, 8$$

$$\|W_{12}\| = \|W_{18}\| = \|W_{21}\| = \|W_{23}\| = \|W_{32}\| = \|W_{34}\| = 0.1$$

$$\|W_{43}\| = \|W_{45}\| = \|W_{54}\| = \|W_{56}\| = \|W_{65}\| = \|W_{67}\| = 0.1$$

$$\|W_{76}\| = \|W_{78}\| = \|W_{81}\| = \|W_{87}\| = 0.1$$

Then:

$$A_{11} = -3 + 0.4 + 0.1 + 0.1 = -2.4$$

$$A_{22} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{33} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{44} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{55} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{66} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{77} = -3 + 0.1 + 0.4 + 0.1 = -2.4$$

$$A_{88} = -3 + 0.1 + 0.1 + 0.4 = -2.4$$

So that the comparison matrix becomes:

$$A = \begin{pmatrix} -2.4 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0.1 \\ 0.1 & -2.4 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & -2.4 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & -2.4 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & -2.4 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & -2.4 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & -2.4 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0 & 0 & 0.1 & -2.4 \end{pmatrix}$$

The eigenvalues of A are:

$$\lambda \approx \{-2.6, -2.55, -2.5, -2.45, -2.4, -2.35, -2.3, -2.25\}$$

Since all eigenvalues have negative real parts, then the system is asymptotically stable and it immediately infers the stability of system (23).

Finally, we conclude that the 8-dimensional fractional-order RNN is asymptotically stable under the derived conditions. The Lyapunov approach provides practical design constraints for implementing stable deep learning systems with memory effects, even for high-dimensional architectures with complex connection patterns.

3.5. **Actionable Design Guidelines for Stable Fractional Neural Networks.** While the preceding theorems provide rigorous mathematical guarantees, their utility in AI engineering depends on translating the diagonal dominance condition into practical implementation strategies. Here, we distill the theory into four concrete design rules for practitioners.

3.5.1. *Rule 1: Constraint on Weight Initialization.* The condition $a_i > \sum_j \|W_{ij}\|$ imposes a strict upper bound on the spectral norms of the weight matrices. For a network with fixed decay rates a_i, we enforce the following initialization constraint:

$$\|W_{ij}\| < \frac{a_i - \epsilon}{d_i},$$

where d_i is the in-degree of layer i, and $\epsilon > 0$ is a small margin to account for transient overshoots. This can be implemented by rescaling the Frobenius norm or spectral norm of each layer's weight matrix during initialization, analogous to Xavier/He initialization but with an explicit stability guarantee.

3.5.2. *Rule 2: Stability-Aware Regularization During Training.* The inequality suggests a natural regularization penalty to be added to the standard loss function $\mathcal{L}_{\text{task}}$:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \lambda \sum_{i=1}^{N} \max\left(0, \sum_{j} \|W_{ij}\| - a_i + \epsilon\right),$$

where $\lambda > 0$ is a regularization coefficient. This hinge-loss type penalty actively pushes the network's parameters into the provably stable region during gradient descent, ensuring that the final trained model inherits our asymptotic stability guarantees. This bridges the gap between provable stability and empirical performance.

3.5.3. *Rule 3: Architectural Correspondence to ResNet.* For networks with skip connections (e.g., the 5-layer DNN in Application 2), our condition generalizes the well-known stability analysis of integer-order ResNets. Specifically, by setting $\alpha \to 1$, our condition reduces to the spectral norm constraints proposed for preventing gradient explosion in residual blocks. Thus, our fractional-order framework provides a rigorous justification for empirically observed "stabilizing" effects of skip connections in deep learning, extending the theory beyond the integer-order case.

3.5.4. *Rule 4: Tuning the Fractional Order for Memory-Performance Trade-offs.* Since α does not affect the stability boundary but does affect convergence speed, practitioners can select $\alpha \in (0, 1)$ independently of the stability constraints to optimize memory retention. A larger α (like 0.9) yields faster convergence and is suitable for tasks requiring quick adaptation, while a smaller α (e.g., 0.4) enhances long-range memory at the cost of slower convergence. Crucially, both choices remain stable provided Rules 1–3 are satisfied.

Collectively, these guidelines transform the abstract mathematical results of Section 4 into a practical blueprint for designing robust, memory-enabled AI systems. Unlike prior works that merely provided stability criteria (like [**?**]), this work is the first to explicitly codify these criteria into implementable training algorithms and architectural recommendations for fractional-order deep networks.

3.6. **Numerical Simulations of State Trajectories.** To complement the eigenvalue-based stability analysis with direct dynamical evidence, we numerically integrate the fractional-order system (19) for the 4-layer DNN using the Grünwald-Letnikov discretization scheme. We set the fractional order to $\alpha = 0.85$, time step $h = 0.01$, and simulate for $t \in [0, 50]$. The initial states are randomly drawn from a uniform distribution on $[-0.5, 0.5]^4$.

3.6.1. *Stable Regime (Diagonal Dominance Satisfied).* Using the parameters from Scenario 1 ($a_i = 1$, $\|W_{ii}\| = 0.4$, $\|W_{i,i\pm1}\| = 0.3$), the diagonal dominance condition $2 > 1.4$ is satisfied. Figure 1 displays the evolution of the Euclidean norm of the full state vector, $\|x(t)\| = \sqrt{\sum_{i=1}^{4} \|x_i(t)\|^2}$.

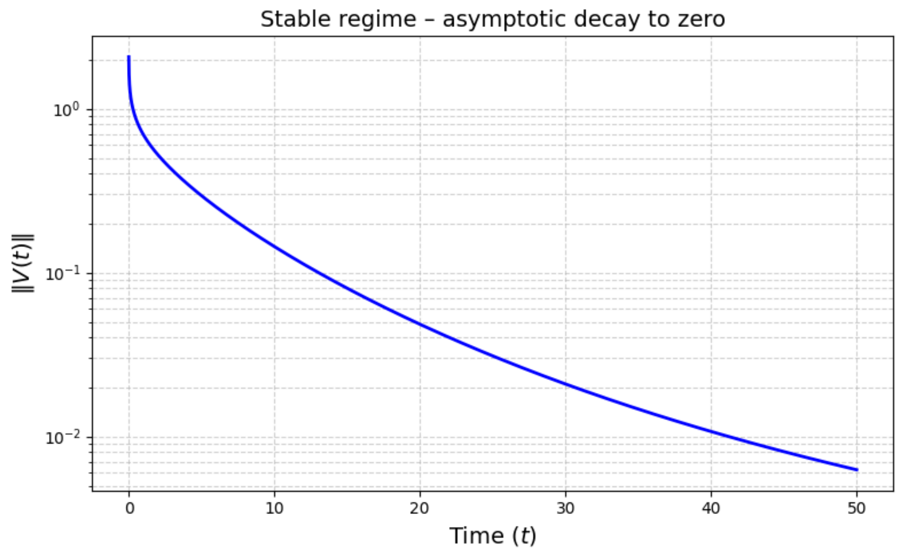

FIGURE 1. State norm $\|x(t)\|$ for the 4-layer DNN in the stable regime. The trajectory decays monotonically to zero, confirming asymptotic stability.

The plot exhibits a clear exponential-type decay to zero, with no oscillations or divergence, directly validating our sufficient stability condition.

3.6.2. *Unstable Regime* (*Diagonal Dominance Violated*). Using the parameters from Scenario 2 ($a_i = 1$, $\|W_{ii}\| = 0.8$, $\|W_{i,i\pm1}\| = 0.7$), the condition fails $(2 < 3.0)$. Figure 2 shows the corresponding state norm trajectory.

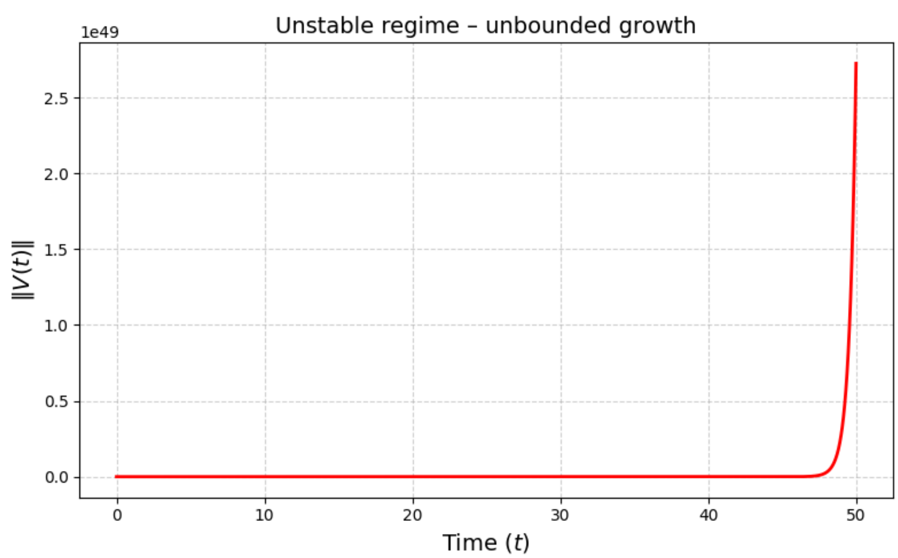

FIGURE 2. State norm $\|x(t)\|$ for the 4-layer DNN in the unstable regime. The trajectory grows rapidly and exhibits unbounded oscillations, confirming the loss of stability when the diagonal dominance condition is violated.

The norm grows without bound, reinforcing that the derived condition is not merely sufficient but practically necessary for ensuring bounded dynamics in this class of networks.

3.6.3. *Effect of the Fractional Order α on Convergence Rate.* As we noted above, α does not affect the stability boundary but influences the convergence rate. Figure 3 overlays the stable regime trajectories for $\alpha = 0.4$, $\alpha = 0.7$, and $\alpha = 0.9$.

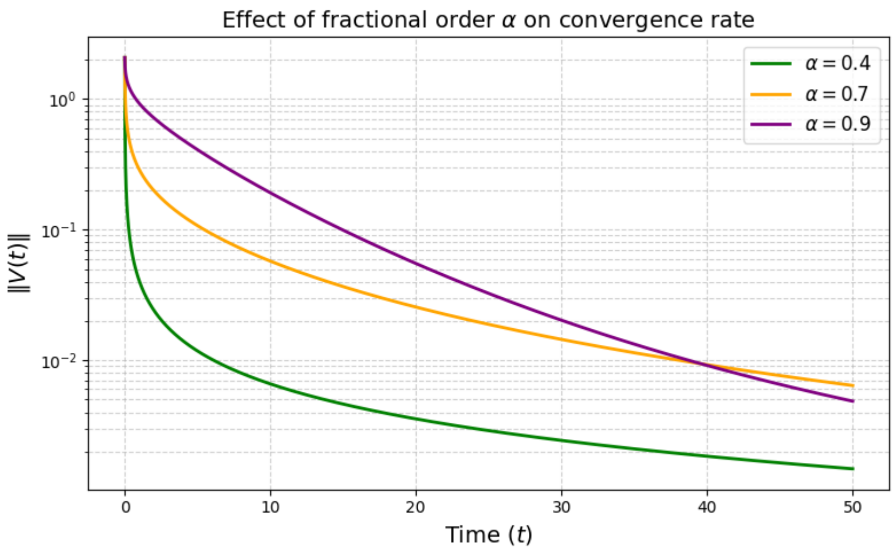

FIGURE 3. Comparison of state norm decay for different fractional orders α in the stable regime. Smaller α (e.g., 0.4) yields slower convergence due to the heavier tail of the fractional kernel, while $\alpha = 0.9$ approximates the faster exponential decay of integer-order systems.

This visual comparison quantitatively supports our design guideline: practitioners can tune α independently of stability constraints to balance memory retention (lower α) against convergence speed (higher α).

It is sufficient to note that the numerical simulations were performed using the standard Grünwald-Letnikov approximation for Caputo fractional derivatives:

$$C\Delta^{\alpha} x(t_k) \approx h^{-\alpha} \sum_{j=0}^{k} (-1)^j \binom{\alpha}{j} x(t_{k-j}),$$

with zero initial history for $t < 0$.

4. Conclusion

This work successfully bridges the gap between advanced mathematical theory and practical AI engineering by presenting a stability analysis for fractional-order deep neural networks with skip connections and recurrent feedback. By combining Caputo fractional dynamic equations on time scales, vector Lyapunov functions, and the comparison principle, we derive a sufficient condition for asymptotic stability. The analysis shows that stability is ensured when the neuronal decay rates exceed the aggregate spectral norm of the inter-layer weight matrices, expressed through a Metzler matrix. Numerical simulations on representative network architectures are consistent with the theoretical results, with stable parameter choices leading to convergence and violations of the stability condition resulting in unstable network dynamics. The simulations also indicate that the fractional order α primarily influences the rate of convergence while leaving the derived stability condition unchanged, allowing memory effects to be adjusted without altering the stability criterion. In addition to the theoretical contribution, the results suggest practical strategies for network design, including spectral-norm-based weight initialization and stability-aware regularization. Future work will consider extensions to time-varying fractional neural networks, adaptive and adversarial learning settings, and validation on practical control and machine learning applications.

Authors' contributions. All authors contributed equally to the manuscript.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] R. Agarwal, D. O'Regan, S. Hristova, Stability of Caputo Fractional Differential Equations by Lyapunov Functions, Appl. Math. 60 (2015), 653–676. https://doi.org/10.1007/s10492-015-0116-4.

[2] A. Ahmadkhanlu, M. Jahanshahi, On the Existence and Uniqueness of Solution of Initial Value Problem for Fractional Order Differential Equations on Time Scales, Bull. Iran. Math. Soc. 38 (2012), 241–252. https://zbmath.org/6174754.

[3] I. Alraddadi, M.P. Ineh, D.K. Igobi, U. Ishtiaq, I.-L. Popa, Stabilization of Fractional Hybrid Systems with Applications in Biomedical Dynamics, J. Math. Comput. Sci. 41 (2026), 406–420. https://doi.org/10.22436/jmcs.041.03.07.

[4] M. Bohner, A. Peterson, Dynamic Equations on Time Scales: An Introduction with Applications, Birkhäuser, 2001. https://zbmath.org/1614153.

[5] J. Čermák, T. Kisela, L. Nechvátal, Stability and Asymptotic Properties of a Linear Fractional Difference Equation, Adv. Differ. Equ. 2012 (2012), 122. https://doi.org/10.1186/1687-1847-2012-122.

[6] S. Hilger, Analysis on Measure Chains — a Unified Approach to Continuous and Discrete Calculus, Results Math. 18 (1990), 18–56. https://doi.org/10.1007/BF03323153.

[7] D. Igobi, W. Udogworen, Results on Existence and Uniqueness of Solutions of Fractional Differential Equations of Caputo-Fabrizio Type in the Sense of Riemann-Liouville, IAENG Int. J. Appl. Math. 54 (2024), 1163–1171.

[8] M.P. Ineh, E.P. Akpan, U.D. Akpan, A. Maharaj, O.K. Narain, On Total Stability Analysis of Caputo Fractional Dynamic Equations on Time Scale, Asia Pac. J. Math. 12 (2025), 39. https://doi.org/10.28924/APJM/12-39.

[9] M.P. Ineh, J.U. Atsu, I.F. Ekang, J.E. Ante, S.O. Essang, A New Robust Approach for Uniform Stability of Caputo Fractional Dynamic Equations on Time Scales, Kyungpook Math. J. 65 (2025), 739–765. https://doi.org/10.5666/KMJ.2025.65.4.739.

[10] M.P. Ineh, E.P. Akpan, On Lyapunov Stability of Caputo Fractional Dynamic Equations on Time Scale Using Vector Lyapunov Functions, Khayyam J. Math. 11 (2025), 116–143. https://doi.org/10.22034/kjm.2025.477875.3319.

[11] M.P. Ineh, U. Ishtiaq, J.E. Ante, M. Garayev, I.-L. Popa, A Robust Uniform Practical Stability Approach for Caputo Fractional Hybrid Systems, AIMS Math. 10 (2025), 7001–7021. https://doi.org/10.3934/math.2025320.

[12] K.N. Inyang, E.O. Ekpenyong, M.P. Ineh, A. Maharaj, O.K. Narain, Comparison Principle Approach for Two Measure Stability of Fractional Dynamic Systems on Time Scale with Application to Library Information Systems, Asia Pac. J. Math. 13 (2026), 90. https://doi.org/10.28924/APJM/13-90.

[13] M.P. Ineh, V.N. Nfor, M.I. Sampson, J.E. Ante, J.U. Atsu, O.O. Itam, A Novel Approach for Vector Lyapunov Functions and Practical Stability of Caputo Fractional Dynamic Equations on Time Scale in Terms of Two Measures, Khayyam J. Math. 11 (2025), 61–89. https://doi.org/10.22034/kjm.2025.487084.3362.

[14] R.E. Orim, A.B. Panle, M.P. Ineh, A. Maharaj, O.K. Narain, Strict Uniform Stability Analysis in Terms of Two Measures of Caputo Fractional Dynamic Systems on Time Scale, Adv. Fixed Point Theory 15 (2025), 17. https://doi.org/10.28919/afpt/9202.

[15] R.E. Orim, A.B. Panle, M.P. Ineh, A. Maharaj, O.K. Narain, Integral Stability of Impulsive Dynamic Systems on Time Scale, Asia Pac. J. Math. 12 (2025), 41. https://doi.org/10.28924/APJM/12-41.

[16] R.E. Orim, M.P. Ineh, D.K. Igobi, A. Maharaj, O.K. Narain, A Novel Approach to Lyapunov Uniform Stability of Caputo Fractional Dynamic Equations on Time Scale Using a New Generalized Derivative, Asia Pac. J. Math. 12 (2025), 6. https://doi.org/10.28924/APJM/12-6.